Evidence of widespread natural recombination among field isolates of equine herpesvirus 4 but not among field isolates of equine herpesvirus 1

P. K. Vaz,1 J. Horsington,1 C. A. Hartley,1 G. F. Browning,1 N. P. Ficorilli,2 M. J. Studdert,2 J. R. Gilkerson2 and J. M. Devlin1

1Asia-Pacific Centre for Animal Health, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia
2Centre for Equine Infectious Diseases, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

Recombination in alphaherpesviruses allows evolution to occur in viruses that have an otherwise stable DNA genome with a low rate of nucleotide substitution. High-throughput sequencing of complete viral genomes has recently allowed natural (field) recombination to be studied in a number of different alphaherpesviruses, however, such studies have not been applied to equine herpesvirus 1 (EHV-1) or equine herpesvirus 4 (EHV-4). These two equine alphaherpesviruses are genetically similar, but differ in their pathogenesis and epidemiology. Both cause economically significant disease in horse populations worldwide. This study used high-throughput sequencing to determine the full genome sequences of EHV-1 and EHV-4 isolates (11 and 14 isolates, respectively) from Australian or New Zealand horses. These sequences were then analysed and examined for evidence of recombination. Evidence of widespread recombination was detected in the genomes of the EHV-4 isolates. Only one potential recombination event was detected in the genomes of the EHV-1 isolates, even when the genomes from an additional 11 international EHV-1 isolates were analysed. The results from this study reveal another fundamental difference between the biology of EHV-1 and EHV-4. The results may also be used to help inform the future safe use of attenuated equine herpesvirus vaccines.

INTRODUCTION

In many alphaherpesviruses, recombination is increasingly being recognized as an important mechanism that plays a major role in evolution of viruses with an otherwise stable DNA genome that have low rates of nucleotide substitution (Lee et al., 2013, 2012; Thiry et al., 2005). Recent advances in high-throughput sequencing of complete viral genomes have allowed natural (field) recombination to be studied in a number of different alphaherpesviruses affecting animals and humans, including herpes simplex virus 1 (HSV-1), varicella-zoster virus (VZV) and infectious laryngotracheitis virus (ILTV) (Kolb et al., 2013; Lee et al., 2013; Norberg et al., 2015, 2006; Peters et al., 2006). Such studies provide insights into virus evolution and also allow the risk of recombination to be assessed in the context of attenuated vaccine use. In 2012, outbreaks of severe respiratory disease in poultry in Australia were attributed to natural recombination events between attenuated vaccine strains of ILTV that generated virulent recombinant viruses (Lee et al., 2012). These outbreaks of disease highlight the importance of studying and understanding herpesvirus recombination in order to protect animal health.

Equine herpesvirus 1 (EHV-1) and equine herpesvirus 4 (EHV-4) are closely related alphaherpesviruses that cause economically significant disease in horses worldwide (Allen et al., 2004; Crabb & Studdert, 1995; Telford et al., 1992, 1998). Although EHV-1 and EHV-4 are genetically very similar, there are a number of important differences in their pathogenesis and epidemiology. Infection with
EHV-4 is most commonly associated with upper respiratory tract disease, but is also occasionally associated with abortion (Allen et al., 2004; Patel & Heldens, 2005). Infection with EHV-1 also causes respiratory disease but infection frequently progresses beyond the upper respiratory tract to result in systemic disease, including abortion and myeloencephalitis (Allen et al., 2004; Edington et al., 1991; Patel & Heldens, 2005; Studdert et al., 2003). Sero-epidemiological studies have revealed a high prevalence of antibodies to EHV-4 in horse populations in different countries, including over 99% sero-positivity in mares and foals tested on a large Thoroughbred stud farm in New South Wales, Australia (Gilkerson et al., 1999). Antibodies to EHV-1 are consistently detected at a lower prevalence than antibodies to EHV-4, with large sero-epidemiological study in Australian horses detecting antibodies to EHV-1 in 26% of mares and 11% of foals (Gilkerson et al., 1999).

Both EHV-1 and -4 have linear, double-stranded type 'D' DNA genomic structures that consist of a unique long (UL) and a unique short (US) genome region, with the US region flanked by large inverted repeats (internal repeat short, IRs, and terminal repeat short, TRs) (Telford et al., 1992, 1998). The EHV-1 and EHV-4 genomes are 150 kbp and 146 kbp in length, respectively. Both encode the same 76 homologous genes, with three duplicated genes in EHV-4 and four duplicated genes in EHV-1 within the repeat regions. The level of amino acid sequence identity between corresponding proteins encoded by the two genomes ranges from 55% to 96% (Telford et al., 1998).

This study aimed to use high-throughput sequencing methods to determine the full genome sequences of a collection of diverse EHV-1 and EHV-4 isolates from Australian and New Zealand horses and to examine these sequences for evidence of recombination. We also aimed to assess the genetic diversity of the EHV-1 and EHV-4 isolates and to examine the phylogenetic relationships between the isolates.

**RESULTS**

**Complete genome sequences of 11 Australian EHV-1 isolates**

The full genome sequences of 11 Australian isolates of EHV-1 (Table 1) were determined by mapping against the reference sequence, or by de novo assembly. Sequence alignments from the former method of assembly are

<table>
<thead>
<tr>
<th>Virus ID</th>
<th>Year</th>
<th>Region</th>
<th>Disease association</th>
<th>Site of collection</th>
<th>Reference</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHV-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>438-77</td>
<td>1977</td>
<td>Scone, NSW</td>
<td>Abortion</td>
<td>Fetus</td>
<td>Studdert &amp; Blackney (1979)</td>
<td>KT324734</td>
</tr>
<tr>
<td>717A-82</td>
<td>1982</td>
<td>VIC</td>
<td>Neurological/abortion</td>
<td>Brain or fetus</td>
<td>Studdert et al. (1984)</td>
<td>KT324733</td>
</tr>
<tr>
<td>770-90</td>
<td>1990</td>
<td>SA</td>
<td>Abortion</td>
<td>NA</td>
<td>This study</td>
<td>KT324732</td>
</tr>
<tr>
<td>1029-93</td>
<td>1993</td>
<td>Wakefield, NSW</td>
<td>Abortion</td>
<td>Fetus</td>
<td>This study</td>
<td>KT324731</td>
</tr>
<tr>
<td>1966-02</td>
<td>2002</td>
<td>Seymour, VIC</td>
<td>Abortion</td>
<td>Fetus</td>
<td>This study</td>
<td>KT324729</td>
</tr>
<tr>
<td>2019-02</td>
<td>2002</td>
<td>Clayton, VIC</td>
<td>Abortion</td>
<td>Fetus</td>
<td>This study</td>
<td>KT324728</td>
</tr>
<tr>
<td>2222-03</td>
<td>2003</td>
<td>Scone, NSW</td>
<td>Abortion</td>
<td>Fetus</td>
<td>This study</td>
<td>KT324727</td>
</tr>
<tr>
<td>3038-07</td>
<td>2007</td>
<td>VIC</td>
<td>Neurological</td>
<td>Nasal septum</td>
<td>This study</td>
<td>KT324726</td>
</tr>
<tr>
<td>3045-07</td>
<td>2007</td>
<td>NSW</td>
<td>Abortion</td>
<td>NA</td>
<td>This study</td>
<td>KT324725</td>
</tr>
<tr>
<td>EHV-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER39-67</td>
<td>1967</td>
<td>SA</td>
<td>Respiratory</td>
<td>Respiratory tract</td>
<td>Studdert et al. (1970)</td>
<td>KT324748</td>
</tr>
<tr>
<td>306-74</td>
<td>1974</td>
<td>Flemington, VIC</td>
<td>Respiratory</td>
<td>Nasal swab</td>
<td>Gleeson et al. (1976)</td>
<td>KT324743</td>
</tr>
<tr>
<td>2387-75</td>
<td>1975</td>
<td>Esperance, WA</td>
<td>Abortion</td>
<td>Fetus</td>
<td>Peet et al. (1978)</td>
<td>KT324742</td>
</tr>
<tr>
<td>403-76</td>
<td>1976</td>
<td>Flemington, VIC</td>
<td>Respiratory</td>
<td>Nasal swab</td>
<td>Gleeson et al. (1976)</td>
<td>KT324740</td>
</tr>
<tr>
<td>3407-77</td>
<td>1977</td>
<td>N/A</td>
<td>Respiratory</td>
<td>NA</td>
<td>This study</td>
<td>KT324739</td>
</tr>
<tr>
<td>3409-77</td>
<td>1977</td>
<td>Gosnells, WA</td>
<td>Abortion</td>
<td>Fetus</td>
<td>Peet et al. (1978)</td>
<td>KT324737</td>
</tr>
<tr>
<td>475-78</td>
<td>1978</td>
<td>Mt Pleasant, TAS</td>
<td>Abortion</td>
<td>Fetus</td>
<td>Studdert &amp; Blackney (1979)</td>
<td>KT324747</td>
</tr>
<tr>
<td>960-90</td>
<td>1990</td>
<td>Diggers Rest, VIC</td>
<td>Respiratory</td>
<td>Nasal swab</td>
<td>Varrasso et al. (2001)</td>
<td>KT324741</td>
</tr>
<tr>
<td>1332-99</td>
<td>1999</td>
<td>Flemington, VIC</td>
<td>Respiratory</td>
<td>Nasal swab</td>
<td>Huang Ja et al. (2002)</td>
<td>KT324744</td>
</tr>
<tr>
<td>1546-99</td>
<td>1999</td>
<td>Toolern Vale, VIC</td>
<td>Respiratory</td>
<td>Nasal swab</td>
<td>Huang Ja et al. (2002)</td>
<td>KT324735</td>
</tr>
<tr>
<td>3056-07</td>
<td>2007</td>
<td>TAS</td>
<td>Respiratory</td>
<td>Nasal swab</td>
<td>This study</td>
<td>KT324746</td>
</tr>
</tbody>
</table>

NA, Not available.
shown in Fig. 1. The results from de novo assembly were principally consistent with those produced by mapping against the reference sequence, with variation observed only in regions rich in repeats (Fig. S2). The estimated size of the EHV-4 full-length genomes ranged from 143.16 bp (isolate 3407-77) to 144.16 kbp (isolate 1532-99). Alignment of the 14 EHV-4 genomes revealed a number of sequence differences between isolates. The genetically most similar isolates shared 99.9 % nucleotide sequence identity (isolates ER39-67 and R19-68; isolates ER39-67 and 475-78; isolates 405-76 and 306-74), excluding large sequence gaps. The genetically most diverse isolates shared 98.9 % nucleotide identity (isolates 960-90 and 1532-99). Details of the predicted amino acid differences between isolates are shown in Table S3. Large regions of sequence difference (insertions/deletions, indels) were present in the repeat regions of the genome, between ORFs 64–65 and ORFs 66–67 of the IR/TR, in ORF24 of the UL region and in ORF71 in the US region. Smaller indels were seen in the UL region within ORFs 24 and 61 and in the US within ORFs 70, 71 and 75. The number of SNPs between different isolates of EHV-4 is summarized in Table S4. The average number of SNPs between EHV-4 isolates was 179.

**Recombination**

Phylogenetic recombination networks for EHV-1 and EHV-4 were generated from nucleotide alignments of the complete genome sequences, as well as alignments of the individual UL, US and IR regions, using SplitsTree4 (Huson, 1998). The multiple reticulate networks and
pair-wise homoplasy (Phi) test analyses (Fig. 3) indicate significant historical recombination events between the different EHV-4 isolates in the UL region ($P<0.001$) but not in the IR or US regions. No significant recombination was detected between the 11 EHV-1 Australian/New Zealand isolates sequenced, irrespective of the genomic region analysed (Fig. 4). No significant recombination events were detected when the analyses were expanded to include the genome sequences from 11 additional international EHV-1 isolates (Fig S3).

Detection of recombination breakpoints utilized six different methods: Recombination detection program (RDP), GENECONV, 3Seq, Maximum Chi Square (MaxChi), SiScan and BootScan. These methods were implemented in the program RDP4 (Martin et al., 2015). Analyses were performed using complete genome sequences, as well as alignments of the individual US, UL and IR regions. Consistent with the results from the SplitsTree analysis, evidence of potential recombination events was detected amongst EHV-4 isolates in the UL region using these different methods (Table 2). In addition, evidence of potential recombination events was detected in the US and IR regions of EHV-4 and another potential recombination event was detected in the IR region of EHV-1 (Table 2). These potential recombination events had not been detected by SplitsTree and Phi test analyses.

**DISCUSSION**

The 11 isolates of EHV-1 and 14 isolates of EHV-4 analysed in this study were all collected from Australian or New Zealand horses between 1967 and 2007. In Australia, EHV-4 was first isolated in 1967, whilst the first isolates of EHV-1 were not obtained until 1977, when it is believed EHV-1 was first introduced into Australia. The sequence data generated in this study have substantially increased the number of full genome sequences available for these viruses, particularly EHV-4, as only one EHV-4 genome has been fully sequenced previously (Telford et al., 1998). Evidence of extensive recombination was detected in the genomes of the EHV-4 isolates, but little or no recombination was detected between the genomes of the EHV-1 isolates, irrespective of whether only the genomes from Australian/New Zealand isolates of EHV-1 and -4 were analysed, or whether genomes from an additional 11 international EHV-1 isolates were included in the analysis.
Analysis of the EHV-1 and EHV-4 full genome sequences showed that the genetic diversity amongst EHV-4 isolates was greater than the diversity among EHV-1 isolates. This is consistent with earlier studies that have assessed genetic differences between EHV-1 and EHV-4 isolates using other techniques, including sequencing of selected genes, or digestion with restriction endonucleases (Allen et al., 2004). This greater diversity in EHV-4 isolates, compared with EHV-1 isolates, is also consistent with the hypothesis that the progenitor virus of EHV-1 was transferred from another host to the ancestor of the modern horse more recently than EHV-4 (Crabb & Studdert, 1990) and has therefore had less time to diversify. An alternative hypothesis, arising from the results from this study, may be that the lack of genetic diversity within EHV-1 is a consequence of a low rate of recombination in this virus species, given that recombination can contribute to genetic diversity.

The lower level of genetic diversity in EHV-1, compared with EHV-4, would make recombination more difficult to detect in EHV-1. The total number of SNPs between any two EHV-1 isolates varied from 8 to 153 (excluding repeat sequences) with an average of 96 SNPs between any two isolates. It is possible that EHV-1 recombination events may have occurred but were not able to be detected at the required level of statistical significance, because an insufficient number of SNPs specific for the parental viruses were transferred to the recombinant viruses. However, evidence of frequent, natural recombination has been readily detected in VZV, which also has a highly conserved genome (approximately 30–200 SNP differences between

---

**Fig. 3.** Recombination network trees generated from Australasian EHV-4 nucleotide alignments (excluding sequence repeats) using SplitsTree4. (a) Complete genome sequences, (b) UL region, (c) US region and (d) repeat region. The multiple reticulate networks indicate recombination events between the different isolates. The bar indicates the rate of evolution in sequence substitutions per site. The Phi test for detecting recombination, as implemented in SplitsTree4, was highly significant for the whole genome, and for the UL region, but not for the US region or the repeat region.
any two strains), using similar methods to those in this study (Norberg et al., 2015, 2006; Zell et al., 2012).

Importantly, we assessed recombination in the EHV isolates using a number of different approaches. Simulated and empirical studies have demonstrated the value of using multiple methods for detecting recombination, rather than relying on a single approach (Posada, 2002; Posada & Crandall, 2001). The pair-wise homoplasy test (Phi test) implemented by SplitsTree is a robust and powerful method for detecting the presence/absence of recombination and has been shown to produce a low rate of false-positive results in a wide variety of circumstances (Bruen et al., 2006). This test detected recombination in the UL region of EHV-4, but did not detect recombination in EHV-1. Analysis of recombination breakpoints, using multiple methods within RDP4, produced some similar results, but also detected evidence of other potential recombination events that were not detected by the Phi test, including within the US and IR regions of EHV-4 and one within the IR region of EHV-1. Although it is expected that different methods of detecting recombination will yield different results (Posada, 2002; Posada & Crandall, 2001), the consistent detection of multiple recombination events in EHV-4, but not in EHV-1, provides substantive evidence that recombination is widespread in EHV-4, but is limited or absent in EHV-1.

The presence of widespread recombination in EHV-4, but not in EHV-1, may be due to differences in the epidemiology of the two viruses (Gilkerson et al., 1999), particularly the lower prevalence of EHV-1 infection and hence the lower rate of co-infection of the same animal with two or more EHV-1s, which is the fundamental requirement for

---

**Fig. 4.** Recombination network trees generated from Australasian EHV-1 nucleotide alignments (excluding sequence repeats) using SplitsTree4. (a) Complete genome sequences, (b) UL region, (c) US region and (d) repeat region. The bar indicates the rate of evolution in sequence substitutions per site. The Phi test for detecting recombination, as implemented in SplitsTree4, was not significant for the whole genome, or for any of the individual genome regions.
intraspacific herpesvirus recombination (Thiry et al., 2005). Intrinsic virus-specific factors could also influence the propensity for recombination. Recent studies in HSV-1 have shown that the gene products of UL12 (5’-to-3’ exonuclease) and UL29 (ssDNA-binding protein) appear to function together as a two-component recombinase to promote recombination in infected cells (Weller & Sawitzke, 2014). The homologues of these genes in EHV-1 and EHV-4 (ORFs 50 and 31, respectively) have a high level of identity between the two virus species but their function has not been directly compared. Future studies to assess and compare recombination in EHV-1 and EHV-4 in cell culture are indicated.

Natural recombination between herpesviruses has been recognized as a safety concern for live attenuated herpesvirus vaccines (Lee et al., 2012; Thiry et al., 2005). Although not currently used in Australia, attenuated vaccines are available for use, or are under development, in other countries to help control disease caused by infection with EHV-1 and EHV-4. The results from this study indicate that for attenuated EHV-4 vaccines, the risks of recombination should be considered when designing and implementing vaccination programmes, but that these considerations may be less of a concern for EHV-1 vaccines.

**METHODS**

**Viruses and cells used in this study.** Eleven Australian or New Zealand isolates of EHV-1 and 14 Australian isolates of EHV-4 (Table 1) were selected from our laboratory archives and propagated in cultures of equine fetal kidney cells (Studdert & Gleeson, 1977) for six passages, including three plaque purifications, in order to amplify sufficient virus for whole genome sequencing. Virions were purified by Ficoll gradient centrifugation (Lee et al., 2011). Alternatively, nucleocapsids were purified as described previously (Pignatti et al., 1979). Viral DNA was extracted from purified nucleocapsids using standard phenol/chloroform extraction methods, or from purified virions using the High Pure PCR Template Preparation kit (Roche) according to manufacturer’s instructions.

**High-throughput sequencing and genome assembly.** High-throughput sequencing was performed as described previously (Lee et al., 2011) and the genome sequences for each isolate were assembled using Geneious V6.1.7 (Kearse et al., 2012), by mapping against the reference sequences of EHV-1 (GenBank accession NC_001491.2) or EHV-4 (GenBank accession NC_001844.1). An overview of the sequencing metrics (including mean coverage and read depth across each genome) is provided in Tables S5 (EHV-1) and S6 (EHV-4). The complete genome sequences of the EHV-1 and EHV-4 isolates were deposited in the NCBI GenBank database under the accession numbers listed in Table 1. For comparison, de novo assembly was also performed for two virus isolates (EHV-1 438-77 and EHV-4 405-76) using medium–low sensitivity settings in Geneious V6.1.7. Contig consensus sequences were aligned to the reference sequence to aid scaffold construction.

**Sequence alignment.** Alignments of the complete genome sequences (excluding the TR regions) of the EHV-1 and EHV-4 isolates, as well as the individual UL, US and IR regions, were prepared using the Multiple Alignment with Fast Fourier Transformation (MAFFT) version 7 plugin in Geneious V6.1.7 (Katoh & Standley, 2013). The laboratory isolates 438-77 and 405-76 (Studdert & Blackney, 1979) were used as the reference sequences for EHV-1 and EHV-4 alignments, respectively.

To expand the analysis of EHV-1 isolates, these same analyses were repeated after including publicly available full genome sequence data from an additional 11 EHV-1 isolates. These included five isolates from Japan (00c19, 90c16, 01c1, 89c105 and 89c25; GenBank accession numbers KF644576, KF644566, KF644578, KF644577 and KF644579, respectively), three isolates from the USA (FL06, NY05 and VA02; GenBank accession numbers KF644567, KF644570 and KF644572, respectively) and three isolates from the UK (NMKT04, V592 and Ab4; GenBank accession numbers KF644576, KF644566, KF644578, KF644577 and KF644579, respectively) were included.

**Recombination analyses.** Tandem repeat regions were identified and removed prior to recombination analysis. For this, perfect and
imperfect repeats were identified in the aligned genomes using the Phobos plugin in Geneious V6.1.7 with default settings and score constraints for satellites. All repeat regions identified were deleted in all viruses in the alignment, including in those viruses in which the repeat region may not have been directly identified due to minor sequence variations that affected the repetitive nature of the sequence. In this way, the deletion of repeat regions did not change the alignment of the genomes. These alignments, and the isolated UL, US or internal repeat regions, were then used to detect evidence of interspecific recombination using two recombination detection programs: SplitsTree4 (Huson, 1998) and RDP4 (Martin et al., 2015). Initially, only the genomes of the Australian/New Zealand EHV isolates were analysed. To expand upon the analyses of EHV1 isolates these same analyses were then repeated after including publicly available full genome sequence data from the additional 11 EHV-1 isolates described above.

In SplitsTree4, splits network trees were generated with an uncorrected P characters transformation model, ignoring constant sites. Other models were tested but resulted in no significant topological differences in the generated trees. Statistical analyses of the recombination networks were performed using the Phi test (Bruen et al., 2006) as implemented by SplitsTree4. In RDP4, six different methods were used to assess the sequences for recombination breakpoints: RDP, GENECONV, 3Seq, SiScan, MaxChi and BootScan. Default RDP4 settings were used throughout. Only breakpoints with a Bonferroni-corrected P value < 0.05 were reported. Duplicate breakpoints were omitted.

**ACKNOWLEDGEMENTS**

This work was supported by the Australian Research Council (ARC, DP130103991). J. M. D. is supported by an ARC Future Fellowship (FT140101287). The authors thank J. Cuxson for her assistance.

**REFERENCES**


Sequence and recombination analysis of EHV-1 and EHV-4


